Audio–Vision Substitution for Blind Individuals: Addressing Human Information Processing Capacity Limitations

David J. Brown and Michael J. Proulx

***Abstract*—In this contribution, we consider the factors that influence the information processing capacity of the person using sensory substitution devices, and the influence of how the translated information, here in audio, impacts performance. First, we review aspects of vision substitution by tactile and audio devices, and then we review key theory in human information processing limitations to devise and test use of an audio–vision substitution device, The vOICe, for recognizing visual objects with audio substitution for vision. Participants heard sonifications of two-dimensional (2-D) images and had to match them to alternatives presented either in visual or tactile modalities. To assess whether capacity limits constrain performance, objects were either presented with all information simultaneously (top and bottom as whole objects), or successively (top and bottom of the object one after the other). Performance was superior in the successive trials, indicative of a capacity limit in processing the auditory information. We discuss the implications for training protocols and design to provide a useful accessibility device for blind individuals.**

***Index Terms*—Assistive Technology, Blindness, Wearable Computers, Sensory Substitution.**

I. Introduction

NUMEROUS technologies and interventions have been developed to make the visual world accessible to those with blindness, legally defined as an acuity of 20/200 in the better eye as a measure of the resolution of information one can process in an optician's test [1]. Invasive techniques such as implants provide low resolution imagery by stimulating surviving retinal cells [2]–[5], cortex [6]–[9] or optic nerve [10]. Aside from risks associated with surgical procedures, these methods are expensive, provide a low functional acuity, and require extensive training to re-establish existing, or stimulate new, neural connections.

Non-invasive methods rely on human-centered computing that bring together signal processing and person-centered computing to harness the plasticity of the person's brain to process information usually attributed to the impaired modality via an unimpaired modality. This method is termed sensory substitution with the prosthesis the sensory substitution device (SSD). In general the substituted modality is vision with the substituting modalities touch [11]–[14] or audition [15]–[17]. Sensory substitution works by using signal processing to convert information from one modality to another. In a sense, the person is thought of as a central processor, with some damage limiting the ability to process input from a camera for vision; this limitation is overcome by having a signal processor that translates visual information (pixels in an image) into another functioning input, such as auditory information for hearing or tactile information for touch. Therefore the central processing person still receives the information normally detected visually in an alternative format that preserves some of the information content. In this contribution we consider the factors, such as amount and density of information, that influence the processing capacity of the person using the device, and the influence of how the translated information, here in audio, impacts performance. First we review aspects of vision substitution by tactile and audio devices, and then we review key theories in human information processing limitations to devise and test the use of the device for recognizing visual objects with audio substitution.

Manuscript received September 15, 2015; revised March 01, 2016; accepted March 01, 2016. Date of publication April 01, 2016; date of current version July 19, 2016. This work was supported by the Engineering and Physical Sciences Research Council under Grant EP/J017205/1. The guest editor coordinating the review of this manuscript and approving it for publication was Dr. Diane Joyce Cook.

The authors are with the Crossmodal Cognition Laboratory, Department of Psychology, University of Bath, Bath BA2 7AY, U.K. (e-mail: djbrown-msp@gmail.com; m.j.proulx@bath.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2016.2543678

II. Sensory Substitution Devices for Visual Impairments

Touch and hearing are the two most common senses used to substitute for a visual impairment using a sensory substitution device. Acuity is low for visual-to-tactile (VT) devices due to limitations in both signal processing and in the sensory capacity of a person. The density of touch receptors (representative of up to 400 functional pixels) provide lower sensitivity than vision, and current devices provide only up to 144 pixels of information due to functional display limitations via the tongue, yet this is adequate for tasks such as simple shape recognition, localization and navigation [18]–[20]. Visual-to-auditory (VA) devices exploit the wide frequency resolution of the cochlea and large dynamic range of the auditory nerve to provide a higher theoretical and functional acuity [21], [22]. Signal processing capacity to translate images into sound is rapid and efficient, and display capacities are much higher than tactile devices - with over 11,000 pixels or 'voicels' attainable using The vOICe SSD [21]. The effectiveness of VA devices has been demonstrated for both of the primary facets of visual perception – what and where, object recognition and localization – in sighted (blindfolded),

congenital and late blind users [23]–[27]. The challenges in learning to use the devices are clear in the limited performance in the studies cited here, and much more needs to be known about the capacity limits of the human users so that the devices are better tailored to the person using them, and for the tasks in which they are used. One great benefit of sensory substitution devices is that they can be used flexibly in many scenarios due to the translation of much of the visual information into another format the person can sense.

The signal processing algorithms used in VA SSD's such as The vOICe [17] utilize natural crossmodal correspondences [28]–[31] in the conversion of visual to auditory information and therefore it is unsurprising that naïve (untrained) users perform above chance in object recognition and localization tasks [23], [24], [27]. Of course it would also be of interest to test whether these 'natural' crossmodal correspondences are found in the congenitally blind where a lack of visual experience may hinder these multisensory associations. However, as with all perceptual learning, prolonged exposure to device use should facilitate an increase in performance, and this highlights the importance of developing effective training protocols to enable the human user to take full advantage of the information being provided.

In training, naïve users are often presented with simple 2 dimensional shapes paired with associated soundscapes, to facilitate an understanding of the conversion algorithm. Maximum contrast between object and background is used to reduce the signal-to-noise ratio, reducing background noise which is irrelevant to the object shape. In visual perception the outer lines of the shape are critical in successful recognition particularly if additional features such as a defined texture are absent. This is analogous in The vOICe in which the frequencies and stereo position of the audio signal describe the 'visual' features of the shape. In real time device use the shape outline can be emphasized by use of a Sobel operator which enhances the shape's edge and reduces task-irrelevant noise from 'interior' sonified pixels [17]. This can be replicated onscreen by using outline shapes rather than ones with a filled interior. While these interior pixels may convey additional information such as surface texture or shading they may impinge on shape recognition through the provision of additional noise. This additional information may also impact on successful performance if it breaches processing capacity limits of the human user.

III. Human Information Processing Capacity

Perception involves the extraction of sensory information from the environment via attentional mechanisms that pass on such information to sensory memory and beyond [32]. Here task relevant information is filtered and subjected to higher-order processing for goal directed action, with task-irrelevant data discarded [33]–[40]. Within these early stages there are posited to be capacity limitations in both the duration and number of pieces of information. For example visual short term memory can retain around 3 or 4 pieces of information for around 10 seconds prior to them being subjected to decay and forgotten, while haptic memory shows a similar duration and set size [41]–[44]. Capacity limitations, or bottlenecks, in sensory perception have been demonstrated in tasks such as the attentional blink, in which a second target may not be fully processed if presented within a certain temporal interval of a preceding stimulus [45], [46], and the psychological refractory period in which the time to make a correct response to one stimulus delays processing on a second [47].

The selection of task-relevant information is modulated by the amount of attentional resources dedicated to processing task-relevant stimuli. Initial attentional models of perception were based on 'early' and 'late' selection. Broadbent [48] and Treisman [49], [50] developed the first models which posited that filtering occurs early in the process based on low-level features such as shape, color and pitch. Conversely, late selection models [51], [52] proposed that capacity is unlimited and all sensory information is automatically attended to equally until higher-order semantic coding selects task-relevant information. As empirical support was found for both models [50], [53], [54], Lavie proposed that the contrasting results on the locus of selection could be explained by perceptual load [55], [56].

The perceptual load theory conceives of perception as a limited-capacity process, as in early selection models, but which proceeds automatically, as in late selection models, until resources are utilized. The locus of attention is therefore modulated by the perceptual load of the task. When the perceptual load of the task is high, processing is dedicated to task-relevant information with task-irrelevant information discarded. Conversely, if perceptual load is low, processing, not exhausted by task-relevant information, permits processing of task-irrelevant information. Thus an increase in perceptual load in task-relevant processing should reduce the extent of interference from irrelevant stimuli [57], [58]. Evidence to support the model has been shown in visual perception demonstrating both inattentional blindness [59] and inattentional deafness [60], [61]. Within other modalities the evidence is less clear. While Santangelo and colleagues (2007) found peripheral auditory cuing effects reduced when the listener was directed to a central auditory stream, (offering support for the perceptual load theory in audition) [62], Murphy failed to find any support for this [63].

If there are capacity limits in sensory substitution then logically this would translate to problems in object recognition. For example, if attentional resources were driven to the high frequencies of the soundscape, and a capacity limit encountered, low frequency data representing the bottom half of the shape would be discarded. In practical terms, a filled circle could be perceived as an upwards facing semi-circle. Therefore it is of interest to test whether such capacity limits exist and, if so, how this can be remedied in training protocols. Unfortunately most perceptual load paradigms are designed to assess normal vision, and utilize the number of distractors and reaction times as dependent measures. In basic training on SSD's accuracy is of most value and it is questionable whether there is actually redundant 'distractor' information, although this is implied by Brown and colleagues [64]. Ideally a paradigm which assesses capacity limits relative to the density of information, uses accuracy as a dependent measure, and could be applied in basic training would be used.

An alternative paradigm, developed to test processing in visual search tasks and using accuracy as a dependent measure, offers a framework for assessing capacity limits in sonified object recognition [65], [66]: the simultaneous versus successive processing task (SIM/SUCC). In the standard SIM/SUCC paradigm, 16 objects are presented on screen either all at the same time (SIMultaneously) or as two SUCCessive eight item displays. If there is a limit to processing capacity then performance on the SUCC condition, where attention is focused on only half of the items in one instance, should be superior to the SIM condition where attention has to be spread over the entire item set. Numerous unimodal studies have used this design to test for limits in capacity in various perceptual tasks such as, visual search, mirror symmetry, perceptual surface completion, attentional blink, and 2D and 3D object shape perception [67]–[70] with capacity limits dependent on task. Using this SIM/SUCC framework we developed a paradigm in which the total information in the signal was presented either simultaneously, or in two successive frames containing half the signal in each.

A final consideration was the sensory modality of the objects used to report recognition of the auditory signal. All participants in the present study were sighted, to demonstrate proof of concept, giving us the option of a visual-to-auditory match. The assumption is that the familiarity of object recognition in the visual modality should facilitate superior performance compared to touch – sighted people evaluate object shape more frequently with the eyes than hands. As visually impaired users have already demonstrated effective device use in object recognition, transferal of results based on capacity limitations should be applicable to such populations. In light of this the experiment was also repeated with a subset of listeners for tactile object matching as this is clearly important for applying the work to those without any visual experience. (though the device could in principle be used for sensory augmentation as well). Based on the literature reviewed here, we make two hypotheses: First, using an established algorithm for the sonification of visual information, presentation of information in the SIM condition will elicit inferior results in the recognition of sonified objects compared to the SUCC condition, implying potential capacity limits; and second, due to this increased load there would be slower reaction times in the SIM condition for both modalities of matching (visual and tactile).

IV. Methods

A. Listeners

A total of 40 (28 female) undergraduate and postgraduate students from Queen Mary University of London and the University of Bath were recruited via email mailing lists. Age ranged between 18 and 31 years with a mean (M) age of 20.63 and a standard deviation (SD) of 2.52. All listeners reported normal or corrected vision and normal hearing with 32 self-reporting as right handed. The study was approved by the Queen Mary University of London Ethics Committee (REC/2009) and the University of Bath Psychology Ethics Committee (#13-204) with all listeners giving written consent prior to commencement of the study. Remuneration was £7 for just the visual task or £12 for completion of visual and tactile sessions.

B. Materials

Visual images for sonification and tactile object creation were obtained from the EST 80 image set (Max Planck Institute, Germany) and Clipart. Stimulus sonification used The vOICe image sonification feature at default settings, and Adobe Audition 3.0. Stimulus presentation was via E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) on a Windows 7 desktop PC. Auditory signals were listened to on Sennheiser HD555 headphones. The blindfold was the Mindfold (Mindfold Inc. Tucson, AZ).

C. Audio-Vision Substitution Signal Processing: The vOICe

The vOICe consists of three hardware components (sensor, processor, and transmitter) and a software algorithm. Visual 'snapshots' at a specified duration are extracted from the environment via a standard webcam. These images are converted to greyscale and then each pixel (174×64) subjected to the principles of the conversion algorithm. The visual brightness of the pixel is coded to auditory amplitude with white pixels at maximum volume then descending through 15 shades of grey to silence for black pixels. The visual elevation of the pixel is coded to auditory frequency on a loglinear scale from 500 Hz (low elevation) to 5000 Hz (high elevation). Pixel position on the x-axis is coded in two ways. At default settings the device scans across the image from left-to-right every 1000 ms with pixels to the left being heard early in the time scan. The device also uses a stereo scan with pixels to the left being heard predominantly in the left headphone. All sonified pixels in a column play concurrently with the 174 columns presented successively across the duration of the scan to give the final sonification of the image. This signal is transmitted back to the user via standard stereo headphones. The device has a number of toggle settings which can be used to manipulate the conversion principles to the user's preferences. For example, the coding of brightness to auditory volume can be switched rendering black pixels as maximum volume, or the time scan duration can be doubled (2000 ms) or reduced down to 125 ms scans.

D. Stimulus Design

White images on a black background were sonified using The vOICe's sonification feature at default settings (1 second scan, normal contrast). Each soundscape's total duration (x axis) was 1000 ms with a total frequency range (y axis) of 500-5000 Hz on a loglinear scale. Bitmap images from The vOICe sonification (keeping relative dimensions) were printed and used as templates for the 5 mm foam board tactile shapes. The foam board cut outs were attached to a background card. Therefore all images presented on screen, tactile objects and associated sonifications were dimensionally consistent.

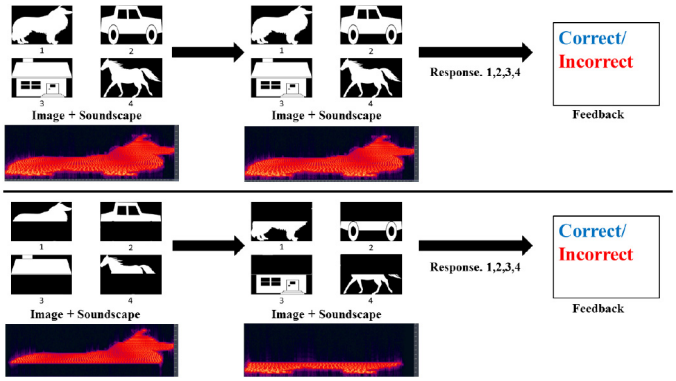

Fig. 1. Example trials for the SIM (top) and SUCC (bottom) conditions. The top panels for each trial display the visual information on screen (or haptic display) for each presentation. The spectrographs below illustrate the energy distribution over time (1000 ms) on the (x-axis) and frequency (500-5000 Hz) on the loglinear scale of the conversion algorithm (y-axis) for the auditory soundscapes.

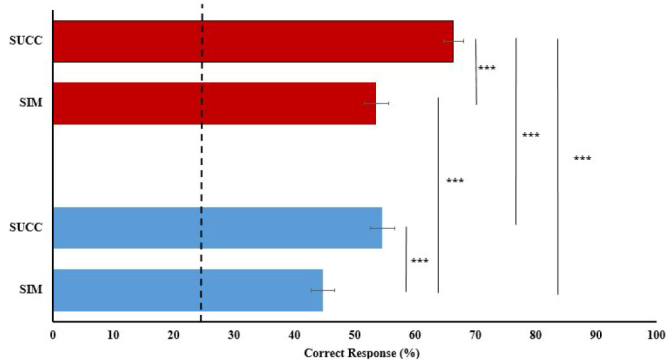

Fig. 2. Object recognition accuracy for the SIM and SUCC conditions for visual-to-auditory (Blue) and tactile-to-auditory (Red) matching tasks. Snificant contrasts are shown by contrast bars (***, $p < .001$) with chance level performance signified by the dotted line. Error Bars represent ± 1 SEM.

Stimuli for the SUCC condition were made by obscuring half the digital image with a black oblong, with the top or bottom edge on the y axis midpoint. Sonifications were made of these 'half' objects. In the tactile matching task card 'masks' were used to obscure the top and bottom of the full tactile objects.

E. Procedure

Listeners watched a PowerPoint presentation describing how The vOICe algorithm converts images to sound including audio-visual explanations of the conversion principles and eight sample shapes – (none from the test set). The second section of the presentation explained the experimental procedure with four example trials.

1) Visual Matching Task (VMT): Fig. 1 shows an example trial from the VMT. For each trial the listener was presented with a four alternative forced choice procedure (4AFC) visual/soundscape association task. Listeners viewed four numbered images on the PC monitor while listening to 1000 ms duration soundscapes, each repeated twice with a 500 ms interstimulus gap. In the SIM condition each of the two soundscapes and four images were of the 'full' objects. For the SUCC condition the soundscape and images were presented successively one 'half' at a time, for example, the top half of the image and audio frequencies followed by the bottom half of the image and audio frequencies. The order of presentation (top or bottom half first) was counterbalanced across participants. The listener's task was to indicate which image the soundscape had been created from by responding 1-4 on the keyboard. Soundscapes and images were repeated twice by default. Each series of 32 randomized trials constituted a block with 3 successive blocks per condition (SIM, SUCC). While accuracy was stressed as the primary objective response times (RT) were also measured from offset of final soundscape (not including self-initiated repeats) to keyboard response. Accuracy feedback was given via a post-trial auditory tone indicating a correct response.

2) Tactile Matching Task (TMT): The basic procedure was similar to the VMT except four tactile, rather than visual, objects were presented to the blindfolded listener to explore haptically while listening to soundscapes. Verbal responses 1-4 were directly inputted by the experimenter who gave tactile accuracy feedback (a tap on the shoulder for correct). For the SUCC condition a card mask was used to obscure the irrelevant half of the tactile object. Due to the much longer trial time (set up and response) the TMT consisted of 2×32 trial blocks per condition. RT response was recorded by the experimenter immediately on verbal response.

To account for order effects the presentation of SIM and SUCC conditions was counterbalanced across participants for both the VMT and TMT. However, all listeners undertook the VMT prior to the TMT.

TABLE I
MEAN ACCURACY AND RESPONES TIMES FOR SIM AND SUCC CONDITIONS IN BOTH THE VISUAL AND TACTILE MATCHING TASKS

	Accuracy (%)	SD	Response Time (ms)	SD
Visual				
SIM	44.77	12.29	3362	1502
SUCC	54.61	13.05	2818	1666
Tactile				
SIM	53.61	8.28	16007	3181
SUCC	66.36	6.51	13659	1991

V. RESULTS

Fig. 2 and Table I show the results for the visual matching task in which listeners were required to match visual objects presented on screen with the heard soundscape in a four alternate forced choice procedure (4AFC).

Initial analysis on the two orders of presentation within the SUCC condition showed that while performance was slightly better when the bottom half ($M = 55.39\%, SD = 13.27$) rather than the top half ($M = 53.83\%, SD = 13.13$) of the stimulus was presented first this was not statistically significant ($t(38) = 0.374, p = .710, d = 0.12$) and therefore for subsequent analysis all successive data will be grouped as SUCC.

Our primary aim was to assess whether the type of presentation facilitated higher levels of object recognition and indeed this was shown to be the case. Both the SIM ($M = 44.77\%$, $SD = 12.39$) condition ($t(39) = 10.089$, $p < .0005, d = 1.60$) and SUCC ($M = 54.61, SD = 13.05$)

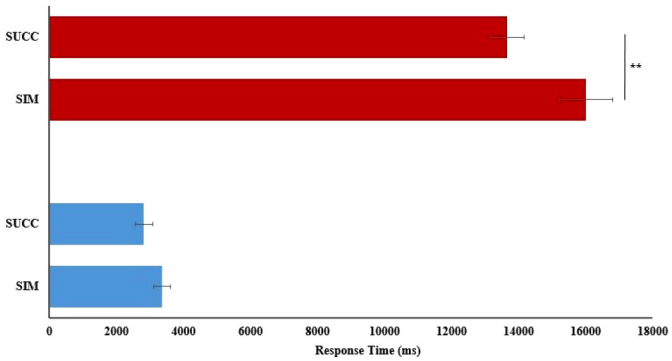

Fig. 3. Response times for the SIM and SUCC conditions for visual-to-auditory (Blue) and tactile-to-auditory (Red) matching tasks. Significant differences are shown by contrast bars+ (**, $p < .01$). Error Bars represent ± 1 SEM. + Note. Due to magnitude of difference in response times for the visual and haptic stimuli contrast bars are only shown within modality. All differences between visual and haptic are significant at $< .0005$.

condition $(t(39) = 14.345, p < .0005, d = 2.27)$ demonstrated object recognition beyond the chance level (25%) illustrating an advantage to using the device irrespective of condition. Comparison of the two types of presentation (SIM + SUCC) clearly showed that object recognition for the SUCC condition was significantly better than for the SIM condition $(t(39) = 6.195), p < .0005, d = 0.98)$.

Next we considered whether the superior performance for the SUCC condition was subject to a speed accuracy trade off, that is, was the higher accuracy associated with longer processing and response times? Results for response times are shown in Fig. 3 and Table I. Analysis in fact showed the opposite in that not only was accuracy higher for this condition but also that the response time for the SUCC condition $(M = 2818 \text{ ms}, SD = 1666)$ was shorter than for the SIM condition $(M = 3362 \text{ ms}, SD = 1502)$ condition although this didn't quite reach significance $(t(39) = 1.956, p = .058, d = 0.31)$.

For the visual matching task the results clearly demonstrate an advantage in object recognition when the total stimulus information was split and presented successively but would this be applicable in a modality relevant to the target user group, the visually impaired? To test this, 18 listeners also participated in the tactile matching task. For this the procedure was similar except that rather than 4 visual objects the listeners were blindfolded and presented with 4 tactile objects. Of the 18 listeners 2 returned incomplete data and so analysis was conducted on 16 listeners data only.

Fig. 2 and Table I show the accuracy and response times for the tactile matching conditions. Again considering the two presentation orders in the SUCC condition performance was better when the top half $(M = 68.65\%, SD = 8.36)$ rather than the bottom half $(M = 64.06\%, SD = 3.01)$ was presented first although as this again was not significant $(t(14) = 1.461, p = .166, d = 0.78)$ the SUCC condition was grouped.

Compared to chance level (25%) for the 4AFC both the SIM $(M = 53.61\%, SD = 8.28)$ condition $(t(15) = 13.819, p < .0005, d = 3.46)$ and SUCC $(M = 66.36\%, SD = 6.51)$ condition $(t(15) = 25.395, p < .0005, d = 6.35)$ were significantly better than expected for the chance level (25%) for the 4AFC. Comparison of the two presentation types again showed that accuracy in the SUCC condition was significantly higher than for the SIM condition $(t(15) = 14.058, p < .0005, d = 3.51)$ with response times, displayed in Fig. 3 and Table I, for the former $(M = 13659 \text{ ms}, SD = 1991)$ again significantly quicker than for the SIM $(M = 16007 \text{ ms}, SD = 3181)$ condition $(t(15) = 2.891, p = .011, d = 0.72)$.

While the focus of the study was on the individual matching tasks all accuracy data was also entered into an ANOVA showing an omnibus main effect $(F(3, 45) = 83.433, p < .0005, \eta p^2 = .848)$. Bonferroni corrected contrast between the four groups showed that the visual SIM condition, aside from being inferior to the visual SUCC also showed worse recognition than for the tactile SIM $(MD = 14.941, 95\%CI [9.83, 20.05], p < .0005)$ and tactile SUCC $(MD = 27.687, 95\%CI [22.40, 32.98], p < .0005)$ conditions. The visual SUCC condition however was only significantly inferior to the tactile SUCC $(MD = 15.088, 95\%CI [9.10, 21.08], p < .0005)$ condition and not the tactile SIM. Considering the vast disparity in response times between the tactile and visual conditions, due to methodology, there seemed little point in contrasting the response times across modality of input.

VI. Discussion

The success of sensory substitution devices is dependent on both the technical specifics of the device and an understanding of how the brain processes multisensory information. Naturally these exist in a feedback loop with a greater understanding of the brain processes informing technological advancements and testing using these new devices giving further insights into multisensory processing. However, consideration should also be given on how to factor in any processing limitations to training protocols for sensory substitution devices and whether specific methodologies can account for these.

The results appear to indicate that there are processing capacity limitations in object recognition using the sensory substitution device in that recognition is superior if the load is halved across presentations. However, we should approach the results with caution as they may not indicate the actual limits but instead may be resultant of the methodology. If processing capacity is limited by the number of sonified pixels then it doesn't necessitate that every split object would be better recognized than a full one. For example, one of the more complex objects, regarding the number of sonified pixels, 'House' consisted of 67.5% white sonified pixels whereas for the 'Diagonal Line' only 9.75% of pixels were sonified. Each frame in the split 'House' sonification would still consist of three times the sonified pixels as the full 'Diagonal Line' and therefore should be more susceptible to any capacity limits, that is, 'Diagonal Line' SIM would outperform 'House' SUCC.

The SIM/SUCC paradigm in visual search [66] uses a number of distractors as a metric in assessing capacity limits but it is questionable whether any information in the sonifications used in the present experiment can be classed as a distractor. The algorithm renders all visual features in the image to an auditory signal but naturally shows no preference for information salient to object recognition, that is, all sonified information at source can be deemed 'useful'. With task-relevant information

extracted from the source signal by lower and higher order processing in the brain selection of salient object features, rather than capacity limits, may be influential in performance. Brown and colleagues [64], in an object recognition paradigm using The vOICe algorithm and the same visual images used here, tested the minimum density of sonified information required for object recognition. Recognition of degraded sonifications created from images at varying pixel resolutions (e.g. 32×32 down to 4×4) displayed a threshold of 8×8 pixels, below which recognition was at chance levels, but with no significant improvement for higher resolution sonifications. The implications being that crude low resolution features, rather than detailed information, were important in object recognition. Applying this to the present experiment, even if capacity limits were exceeded in high density SIM presentations there should still be sufficient information for object recognition.

The more likely explanation for the results would appear to lie in the methodology of the task procedure, specifically how it guides the user to object features. While we are unaware of any auditory attention bias up or down, splitting the object signal directs the user to these frequencies independently, analogous to work on capacity limits in vision [66], [68]. If features such as texture and shading are removed then shape is critical in object recognition, particularly the outer edges. Within the sonification paradigm, how the highest frequencies develop over the duration of the time scan gives an indication of the shape of the top edge of the object. Conversely development of the low frequencies informs on the bottom edge. Presented with the full object the listener has a choice as to where attentional processes are directed; top edge, bottom edge, or both simultaneously, and thus misidentification of objects could be a function of where attention is directed. The SUCC presentation, in which the object is split based on auditory frequency, directly draws the listeners' attention to high and low frequency components independently in time, allowing a more thorough evaluation of the object features separately, and thus with perhaps greater fidelity for each part.

While the study hints at potential capacity limits in sensory substitution further research should assess whether this is consistent across other features of the stimulus. For example, in the present study objects were split on the y-axis (frequency) but consistent on the x-axis (duration). The sonified top and bottom edges of the shape are thus heard independently of each other in the SUCC condition. Reversing the plane on which the object is divided (duration) while splitting the information load would result in concurrent presentation of the frequencies representing the top and bottom edges of the object, as in the SIM condition. Similarly, objects could be split by visual brightness/auditory amplitude in a comparable design. Analysis of such results would give further indication as to whether it is processing capacities or directed attention that facilitates superior performance within the paradigm. The application of the present methodology in training does show potential however. At early stages of use guiding attention to object features may help facilitate a better understanding of the image to sound conversion process and allow the user to develop effective strategies for 'real-time' device use. Outside of the lab context, technological filtering of the source signal to present high and low frequency frames successively could be utilized but it seems more logical and resource effective to let the brain do the work. Strategies where the user attends to the high and low frequency components of the signal should be proposed and tested and, if successful, factored into training regimes for learning to use sensory substitution devices.

VII. Conclusion

In both the visual and tactile matching tasks trials in which the information load was split across two successive presentations displayed both superior recognition and quicker response times compared to when all information was presented in a simultaneous manner. Surprisingly accuracy for the tactile trials was generally higher than for the visual ones although unsurprisingly response times for the latter were significantly quicker.

References

[1] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010," *Brit. J. Ophthalmol.*, vol. 96, no. 5, pp. 614–618, May 2012.

[2] E. Noorsal, K. Sooksood, H. C. Xu, R. Hornig, J. Becker, and M. Ortmanns, "A neural stimulator frontend with high-voltage compliance and programmable pulse shape for epiretinal implants," *IEEE J. Solid-State Circuits*, vol. 47, no. 1, pp. 244–256, Jan. 2012.

[3] M. Keseru *et al.*, "Acute electrical stimulation of the human retina with an epiretinal electrode array," *Acta Ophthalmol.*, vol. 90, no. 1, pp. e1–e8, Feb. 2012.

[4] M. Eickenscheidt, M. Jenkner, R. Thewes, P. Fromherz, and G. Zeck, "Electrical stimulation of retinal neurons in epiretinal and subretinal configuration using a multicapacitor array," *J. Neurophysiol.*, vol. 107, no. 10, pp. 2742–2755, May 2012.

[5] E. Zrenner *et al.*, "Subretinal electronic chips allow blind patients to read letters and combine them to words," *Proc. R. Soc. B Biol. Sci.*, vol. 278, no. 1711, pp. 1489–1497, May 22, 2011.

[6] E. M. Schmidt, M. J. Bak, F. T. Hambrecht, C. V. Kufta, D. K. O'Rourke, and P. Vallabhanath, "Feasibility of a visual prosthesis for the blind based on intracortical microstimulation of the visual cortex," *Brain*, vol. 119, pp. 507–522, Apr. 1996.

[7] R. A. Normann, E. M. Maynard, P. J. Rousche, and D. J. Warren, "A neural interface for a cortical vision prosthesis," *Vis. Res.*, vol. 39, no. 15, pp. 2577–2587, Jul. 1999.

[8] W. H. Dobelle, M. G. Mladejovsky, and J. P. Girvin, "Artifical vision for the blind: Electrical stimulation of visual cortex offers hope for a functional prosthesis," *Science*, vol. 183, no. 4123, pp. 440–444, Feb. 1, 1974.

[9] G. S. Brindley and W. S. Lewin, "The visual sensations produced by electrical stimulation of the medial occipital cortex," *J. Physiol.*, vol. 194, no. 2, pp. 54–5P, Feb. 1968.

[10] C. Veraart, M.-C. Wanet-Defalque, B. Gérard, A. Vanlierde, and J. Delbeke, "Pattern recognition with the optic nerve visual prosthesis," *Artif. Organs*, vol. 27, no. 11, pp. 996–1004, 2003.

[11] A. Arnoldussen and D. C. Fletcher, "Visual perception for the blind: The brainport vision device," *Retinal Phys.*, vol. 9, no. 1, pp. 32–34, 2012.

[12] P. Bach-y-Rita, "Tactile sensory substitution studies," *Ann. New York Acad. Sci.*, vol. 1013, pp. 83–91, May 2004.

[13] P. Bach-y-Rita, C. C. Collins, B. White, F. A. Saunders, L. Scadden, and R. Blomberg, "A tactile vision substitution system," *Amer. J. Optom. Arch. Amer. Acad. Optom.*, vol. 46, no. 2, pp. 109–111, Feb. 1969.

[14] Y. P. Danilov, M. E. Tyler, K. L. Skinner, R. A. Hogle, and P. Bach-y-Rita, "Efficacy of electrotactile vestibular substitution in patients with peripheral and central vestibular loss," *J. Vestibular Res.*, vol. 17, no. 2–3, pp. 119–130, 2007.

[15] S. Abboud, S. Hanassy, S. Levy-Tzedek, S. Maidenbaum, and A. Amedi, "EyeMusic: Introducing a "visual" colorful experience for the blind using auditory sensory substitution," *Restor. Neurol. Neurosci.*, vol. 32, no. 2, pp. 247–257, 2014.

[16] C. Capelle, C. Trullemans, P. Arno, and C. Veraart, "A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 10, pp. 1279–1293, Oct. 1998.

[17] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 2, pp. 112–121, Feb. 1992.

[18] D. R. Chebat, C. Rainville, R. Kupers, and M. Ptito, "Tactile-'visual' acuity of the tongue in early blind individuals," *Neuroreport*, vol. 18, no. 18, pp. 1901–1904, Dec. 3, 2007.

[19] D. R. Chebat, F. C. Schneider, R. Kupers, and M. Ptito, "Navigation with a sensory substitution device in congenitally blind individuals," *Neuroreport*, vol. 22, no. 7, pp. 342–347, May 11, 2011.

[20] M. Ptito, S. M. Moesgaard, A. Gjedde, and R. Kupers, "Cross-modal plasticity revealed by electrotactile stimulation of the tongue in the congenitally blind," *Brain*, vol. 128, pp. 606–614, Mar. 2005.

[21] A. Haigh, D. J. Brown, P. Meijer, and M. J. Proulx, "How well do you see what you hear? The acuity of visual-to-auditory sensory substitution," *Front. Psychol.*, vol. 4, pp. 330, 2013.

[22] E. Striem-Amit, M. Guendelman, and A. Amedi, "'Visual' acuity of the congenitally blind using visual-to-auditory sensory substitution," *Plos One*, vol. 7, no. 3, Mar. 16, 2012.

[23] D. J. Brown, T. Macpherson, and J. Ward, "Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device," *Perception*, vol. 40, no. 9, pp. 1120–1135, 2011.

[24] J. K. Kim and R. J. Zatorre, "Generalized learning of visual-to-auditory substitution in sighted individuals," *Brain Res.*, vol. 1242, pp. 263–275, Nov. 25, 2008.

[25] J. K. Kim and R. J. Zatorre, "Can you hear shapes you touch?," *Exp. Brain Res.*, vol. 202, no. 4, pp. 747–54, May 2010.

[26] C. Poirier, A. De Volder, D. Tranduy, and C. Scheiber, "Pattern recognition using a device substituting audition for vision in blindfolded sighted subjects," *Neuropsychologia*, vol. 45, no. 5, pp. 1108–1121, 2007.

[27] M. J. Proulx, P. Stoerig, E. Ludowig, and I. Knoll, "Seeing 'where' through the ears: Effects of learning-by-doing and long-term sensory deprivation on localization based on image-to-sound substitution," *Plos One*, vol. 3, no. 3, e1840, Mar. 26, 2008.

[28] E. Ben-Artzi and L. E. Marks, "Visual-auditory interaction in speeded classification: Role of stimulus difference," *Percept. Psychophys.*, vol. 57, no. 8, pp. 1151–1162, Nov. 1995.

[29] I. H. Bernstein and B. A. Edelstein, "Effects of some variations in auditory input upon visual choice reaction time," *J. Exp. Psychol.*, vol. 87, no. 2, pp. 241–247, Feb. 1971.

[30] L. E. Marks, "On cross-modal similarity: Auditory visual interactions in speeded discrimination," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 13, no. 3, pp. 384–394, Aug. 1987.

[31] J. C. Stevens and L. E. Marks, "Cross-modality matching of brightness and loudness," *Proc. Natl. Acad. Sci. USA*, vol. 54, no. 2, pp. 407–411, Aug. 1965.

[32] H. E. Pashler, *The Psychology of Attention*. Cambridge, MA, USA: MIT Press, 1998.

[33] S. N. Hajimirza, M. J. Proulx, and E. Izquierdo, "Reading users' minds from their eyes: A Method for implicit image annotation," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 805–815, Jun. 2012.

[34] M. J. Proulx, "Turning on the spotlight: Do attention and luminance contrast affect neuronal responses in the same way?," *J. Neurosci.*, vol. 27, no. 48, pp. 13043–13044, Nov. 28, 2007.

[35] M. J. Proulx, "Bottom-up guidance in visual search for conjunctions," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 33, no. 1, pp. 48–56, Feb. 2007.

[36] M. J. Proulx, "Size matters: Large objects capture attention in visual search," *PLoS One*, vol. 5, no. 12, p. e15293, 2010.

[37] M. J. Proulx, "Individual differences and metacognitive knowledge of visual search strategy," *PLoS One*, vol. 6, no. 10, p. e27043, 2011.

[38] M. J. Proulx and H. E. Egeth, "Target-nontarget similarity modulates stimulus-driven control in visual search," *Psychon. Bull. Rev.*, vol. 13, no. 3, pp. 524–529, Jun. 2006.

[39] M. J. Proulx and M. Green, "Does apparent size capture attention in visual search? evidence from the Muller-Lyer illusion," *J. Vis.*, vol. 11, no. 13, pp. 1–6, Nov. 23, 2011 [Online]. Available: http://www.journalofvision.org/content/11/13/21, doi: 10.1167/11.13.21.

[40] M. J. Proulx and J. T. Serences, "Searching for an oddball: Neural correlates of singleton detection mode in parietal cortex," *J. Neurosci.*, vol. 26, no. 49, pp. 12631–12632, Dec. 6, 2006.

[41] J. C. Bliss, H. D. Crane, P. K. Mansfield, and J. T. Townsend, "Information available in brief tactile presentations," *Percept. Psychophys.*, vol. 1, no. 4, pp. 273–283, 1966.

[42] Y. Jiang, I. R. Olson, and M. M. Chun, "Organization of visual short-term memory," *J. Exp. Psychol. Learn. Memory Cognit.*, vol. 26, no. 3, p. 683, 2000.

[43] S. J. Luck and E. K. Vogel, "The capacity of visual working memory for features and conjunctions," *Nature*, vol. 390, no. 6657, pp. 279–281, Nov. 20, 1997.

[44] H. Pashler, "Familiarity and visual change detection," *Percept. Psychophys.*, vol. 44, no. 4, pp. 369–378, Oct. 1988.

[45] D. Shen and T. A. Mondor, "Effect of distractor sounds on the auditory attentional blink," *Percept. Psychophys.*, vol. 68, no. 2, pp. 228–243, Feb. 2006.

[46] S. Tremblay, F. Vachon, and D. M. Jones, "Attentional and perceptual sources of the auditory attentional blink," *Percept. Psychophys.*, vol. 67, no. 2, pp. 195–208, Feb. 2005.

[47] H. Pashler, "Dual-task interference in simple tasks: Data and theory," *Psychol. Bull.*, vol. 116, no. 2, pp. 220–244, Sep. 1994.

[48] D. E. Broadbent, *Perception and Communication*. New York, NY, USA: Pergamon Press, 1958.

[49] A. Treisman and G. Geffen, "Selective attention: Perception or response?," *Q. J. Exp. Psychol.*, vol. 19, no. 1, pp. 1–17, Feb. 1967.

[50] A. Treisman and J. G. Riley, "Is selective attention selective perception or selective response? A further test," *J. Exp. Psychol.*, vol. 79, no. 1, pp. 27–34, Jan. 1969.

[51] J. A. Deutsch, D. Deutsch, P. H. Lindsay, and A. M. Treisman, "Comments and reply on "Selective attention: Perception or response?," *Q. J. Exp. Psychol.*, vol. 19, no. 4, pp. 362–367, Nov. 1967.

[52] D. A. Norman, "Toward a theory of memory and attention," *Psychol. Rev.*, vol. 75, no. 6, p. 522, 1968.

[53] J. Miller, "Priming is not necessary for selective-attention failures: Semantic effects of unattended, unprimed letters," *Percept. Psychophys.*, vol. 41, no. 5, pp. 419–434, 1987.

[54] C. R. Snyder, "Selection, inspection, and naming in visual search," *J. Exp. Psychol.*, vol. 92, no. 3, p. 428, 1972.

[55] N. Lavie and Y. Tsal, "Perceptual load as a major determinant of the locus of selection in visual attention," *Percept. Psychophys.*, vol. 56, no. 2, pp. 183–197, Aug. 1994.

[56] N. Lavie, "Perceptual load as a necessary condition for selective attention," *J. Exp. Psychol. Human Percept. Perform.*, vol. 21, no. 3, pp. 451–468, Jun. 1995.

[57] N. Lavie, "The role of perceptual load in visual awareness," *Brain Res.*, vol. 1080, no. 1, pp. 91–100, Mar. 29, 2006.

[58] J. S. Macdonald and N. Lavie, "Load induced blindness," *J. Exp. Psychol. Human Percept. Perform.*, vol. 34, no. 5, pp. 1078–1091, Oct. 2008.

[59] U. Cartwright-Finch and N. Lavie, "The role of perceptual load in inattentional blindness," *Cognition*, vol. 102, no. 3, pp. 321–340, Mar. 2007.

[60] J. S. Macdonald and N. Lavie, "Visual perceptual load induces inattentional deafness," *Atten. Percept. Psychophys.*, vol. 73, no. 6, pp. 1780–1789, Aug. 2011.

[61] D. Raveh and N. Lavie, "Load-induced inattentional deafness," *Atten. Percept. Psychophys.*, vol. 77, no. 2, pp. 483–492, Feb. 2015.

[62] V. Santangelo, M. Olivetti Belardinelli, and C. Spence, "The suppression of reflexive visual and auditory orienting when attention is otherwise engaged," *J. Exp. Psychol. Human Percept. Perform.*, vol. 33, no. 1, pp. 137–148, Feb. 2007.

[63] S. Murphy, N. Fraenkel, and P. Dalton, "Perceptual load does not modulate auditory distractor processing," *Cognition*, vol. 129, no. 2, pp. 345–355, Nov. 2013.

[64] D. J. Brown, A. J. Simpson, and M. J. Proulx, "Visual objects in the auditory system in sensory substitution: How much information do we need?," *Multisensory Res.*, vol. 27, no. 5–6, pp. 337–357, 2014.

[65] C. W. Eriksen and T. Spencer, "Rate of information processing in visual perception: Some results and methodological considerations," *J. Exp. Psychol.*, vol. 79, no. 2, pp. 1–16, Feb. 1969.

[66] R. M. Shiffrin and G. T. Gardner, "Visual processing capacity and attentional control," *J. Exp. Psychol.*, vol. 93, no. 1, pp. 72–82, Apr. 1972.

[67] M. Attarha, C. M. Moore, A. Scharff, and J. Palmer, "Evidence of unlimited-capacity surface completion," *J. Exp. Psychol. Human Percept. Perform.*, vol. 40, no. 2, pp. 556–565, Apr. 2014.

[68] L. Huang and H. Pashler, "Attention capacity and task difficulty in visual search," *Cognition*, vol. 94, no. 3, pp. B101–11, Jan. 2005.

[69] L. Huang, H. Pashler, and J. A. Junge, "Are there capacity limitations in symmetry perception?," *Psychon. Bull. Rev.*, vol. 11, no. 5, pp. 862–869, Oct. 2004.

[70] A. Scharff, J. Palmer, and C. M. Moore, "Divided attention limits perception of 3-D object shapes," *J. Vis.*, vol. 13, no. 2, p. 18, 2013.

David J. Brown received the B.Sc. degree in psychology and the M.Res. degree the University of Sussex, Brighton, U.K., and the Ph.D. degree in experimental psychology with Queen Mary University of London, London, U.K.

Presently, he is a Postdoctoral Research Fellow with computer science and a member of the Crossmodal Cognition Laboratory, University of Bath, Bath, U.K. He has published in numerous scientific journals, given talks at national and international conferences, and interviews for a variety of media outlets. His research interests include multisensory perception including perceptual learning in visual-to-auditory sensory substitution, affect through haptic and sonified stimulation, information processing capacity, and general crossmodal correspondences.

Michael J. Proulx received the B.Sc. degree in psychology from Arizona State University, Tempe, AZ, USA, and the M.A. and Ph.D. degrees in psychological and brain sciences from Johns Hopkins University, Baltimore, MD, USA.

He is an Associate Professor (Senior Lecturer) with the Department of Psychology, University of Bath, Bath, U.K. and the Director of the Crossmodal Cognition Laboratory, where he is also affiliated with the Centre for Digital Entertainment, Department of Computer Science. He has authored more than 50 technical papers and book chapters on multisensory integration, attention, ergonomics, auditory-to-visual sensory substitution, synaesthesia, visual search, and eye-tracking for image annotation. He is a Fellow of the Society of Experimental Psychology and Cognitive Science (American Psychological Association).

Dr. Proulx is a member of the Editorial Boards for *Restorative Neurology and Neuroscience* and *PLoS ONE* and a Guest Editor for a special issue of *Neuroscience and Biobehavioral Reviews* and a forthcoming issue of *Multisensory Research.* His international expertise has been acknowledged through awards and invited presentations worldwide, including being honored as a torchbearer for the London 2012 Paralympic Games. He was the recipient of the APA Division of Experimental Psychology New Investigator Award in Human Perception and Performance.